within a genome, and thus would tend to be conserved across individuals. For example, hybridization selections may be made for non-repetitive and single copy sequences. See, e.g., Britten and Kohne (1968) "Repeated Sequences in DNA," Science 161:529-540. On the other hand, it may be desired under certain circumstances to use repeated sequences. For example, where a fingerprint may be used to identify or distinguish different species, or where repetitive sequences may be diagnostic of specific species, repetitive sequences may be desired for inclusion in the fingerprinting probes. In either case, the sequencing capability will greatly assist in the selection of appropriate sequences to be used as probes.

Also as indicated above, various means for constructing an appropriate substrate may involve either mechanical or automated procedures. The standard VLSIPS automated procedure involves synthesizing oligonucleotides or short polymers directly on the substrate. In various other embodiments, it is possible to attach separately synthesized reagents onto the matrix in an ordered array. Other circumstances may lend themselves to transfer a pattern from a petri plate onto a solid substrate. Also, there are methods for site specifically directing collections of reagents to specific locations using unnatural nucleotides or equivalent sorts of targeting molecules.

While a brute force manual transfer process may be utilized sequentially for attaching various samples to successive positions, instrumentation for automating such procedures may also be devised. The automated system for performing such would preferably be relatively easily designed and conceptually easily understood.

XIV. COMMERCIAL APPLICATIONS

A. Sequencing

As indicated above, sequencing may be performed either de novo or as a verification of another sequencing method. The present hybridization technology provides the ability to sequence nucleic acids and polynucleotides de novo, or as a means to verify either the Maxam and Gilbert chemical

sequencing technique or Sanger and Coulson dideoxy- sequencing techniques. The hybridization method is useful to verify sequencing determined by any other sequencing technique and to closely compare two similar sequences, e.g., to identify and locate sequence differences.

Besides polynucleotide sequencing, the present invention also provides means for sequencing other polymers. This includes polypeptides, carbohydrates, synthetic organic polymers, and other polymers. Again, the sequencing may be either verification or de novo.

Of course, sequencing can be very important in many different sorts of environments. For example, it will be useful in determining the genetic sequence of particular markers in various individuals. In addition, polymers may be used as markers or for information containing molecules to encode information. For example, a short polynucleotide sequence may be included in large bulk production samples indicating the manufacturer, date, and location of manufacture of a product. For example, various drugs may be encoded with this information with a small number of molecules in a batch. For example, a pill may have somewhere from 10 to 100 to 1,000 or more very short and small molecules encoding this information. When necessary, this information may be decoded from a sample of the material using a polymerase chain reaction (PCR) or other amplification method. This encoding system may be used to provide the origin of large bulky samples without significantly affecting the properties of those samples. For example, chemical samples may also be encoded by this method thereby providing means for identifying the source and manufacturing details of lots. The origin of bulk hydrocarbon samples may be encoded. Production lots of organic compounds such as benzene or plastics may be encoded with a short molecule polymer. Food stuffs may also be encoded using similar marking molecules. Even toxic waste samples can be encoded determining the source or origin. In this way, proper disposal can be traced or more easily enforced.

Similar sorts of encoding may be provided by fingerprinting-type analysis. Whether the resolution is

absolute or less so, the concept of coding information on molecules such as nucleic acids, which can be amplified and later decoded, may be a very useful and important application.

This technology also provides the ability to include markers for origins of biological materials. For example, a patented animal line may be transformed with a particular unnatural sequence which can be traced back to its origin. With a selection of multiple markers, the likelihood could be negligible that a combination of markers would have independently arisen from a source other than the patented or specifically protected source. This technique may provide a means for tracing the actual origin of particular biological materials. Bacteria, plants, and animals will be subject to marking by such encoding sequences.

B. Fingerprinting

As indicated above, fingerprinting technology may also be used for data encryption. Moreover, fingerprinting allows for significant identification of particular individuals. Where the fingerprinting technology is standardized, and used for identification of large numbers of people, related equipment and peripheral processing will be developed to accompany the underlying technology. For example, specific equipment may be developed for automatically taking a biological sample and generating or amplifying the information molecules within the sample to be used in fingerprinting analysis. Moreover, the fingerprinting substrate may be mass produced using particular types of automatic equipment. Synthetic equipment may produce the entire matrix simultaneously by stepwise synthetic methods as provided by the VLSIPS™ technology. The attachment of specific probes onto a substrate may also be automated, e.g., making use of the caged biotin technology. See, e.g., Barrett et al. (1993) U.S. Pat. No. 5,252,743. As indicated above, there are automated methods for actually generating the matrix and substrate with distinct sequence reagents positionally located at each of the matrix positions. Where such reagents are, e.g., unnatural amino acids, a targeting function may be utilized which does not

interfere with a natural nucleotide functionality.

In addition, peripheral processing may be important and may be dedicated to this specific application. Thus, automated equipment for producing the substrates may be designed, or particular systems which take in a biological sample and output either a computer readout or an encoded instrument, e.g., a card or document which indicates the information and can provide that information to others. An identification having a short magnetic strip with a few million bits may be used to provide individual identification and important medical information useful in a medical emergency.

In fact, data banks may be set up to correlate all of this information of fingerprinting with medical information. This may allow for the determination of correlations between various medical problems and specific DNA sequences. By collating large populations of medical records with genetic information, genetic propensities and genetic susceptibilities to particular medical conditions may be developed. Moreover, with standardization of substrates, the micro encoding data may be also standardized to reproduce the information from a centralized data bank or on an encoding device carried on an individual person. On the other hand, if the fingerprinting procedure is sufficiently quick and routine, every hospital may routinely perform a fingerprinting operation and from that determine many important medical parameters for an individual.

In particular industries, the VLSIPS sequencing, fingerprinting, or mapping technology will be particularly appropriate. As mentioned above, agricultural livestock suppliers may be able to encode and determine whether their particular strains are being used by others. By incorporating particular markers into their genetic stocks, the markers will indicate origin of genetic material. This is applicable to seed producers, livestock producers, and other suppliers of medical or agricultural biological materials.

This may also be useful in identifying individual animals or plants. For example, these markers may be useful in determining whether certain fish return to their original breeding grounds, whether sea turtles always return to their

original birthplaces, or to determine the migration patterns and viability of populations of particular endangered species. It would also provide means for tracking the sources of particular animal products. For example, it might be useful for determining the origins of controlled animal substances such as elephant ivory or particular bird populations whose importation or exportation is controlled.

As indicated above, polymers may be used to encode important information on source and batch and supplier. This is described in greater detail, e.g., "Applications of PCR to industrial problems," (1990) in Chemical and Engineering News 68:145, which is hereby incorporated herein by reference. In fact, the synthetic method can be applied to the storage of enormous amounts of information. Small substrates may encode enormous amounts of information, and its recovery will make use of the inherent replication capacity. For example, on regions of 10 μm x 10 μm, 1 cm^2 has 10^4 regions. In theory, the entire human genome could be attached in 1000 nucleotide segments on a 3 cm^2 surface. Genomes of endangered species may be stored on these substrates.

Fingerprinting may also be used for genetic tracing or for identifying individuals for forensic science purposes. See, e.g., Morris, J. et al. (1989) "Biostatistical Evaluation of Evidence From Continuous Allele Frequency Distribution DNA Probes in Reference to Disputed Paternity and Identity," J. Forensic Science 34:1311-1317, and references provided therein; each of which is hereby incorporated herein by reference.

In addition, the high resolution fingerprinting allows the distinguishability to high resolution of particular samples. As indicated above, new cell classifications may be defined based on combinations of a large number of properties. Similar applications will be found in distinguishing different species of animals or plants. In fact, microbial identification may become dependent on characterization of the genetic content. Tumors or other cells exhibiting abnormal physiology will be detectable by use of the present invention. Also, knowing the genetic fingerprint of a microorganism may provide very useful information on how to treat an infection by

such organism.

Modifications of the fingerprint embodiments may be used to diagnose the condition of the organism. For example, a blood sample is presently used for diagnosing any of a number of different physiological conditions. A multi-dimensional fingerprinting method made available by the present invention could become a routine means for diagnosing an enormous number of physiological features simultaneously. This may revolutionize the practice of medicine in providing information on an enormous number of parameters together at one time. In another way, the genetic predisposition may also revolutionize the practice of medicine providing a physician with the ability to predict the likelihood of particular medical conditions arising at any particular moment. It also provides the ability to apply preventive medicine.

The present invention might also find application in use for screening new drugs and new reagents which may be very important in medical diagnosis or other applications. For example, a description of generating a population of monoclonal antibodies with defined specificities may be very useful for producing various drugs or diagnostic reagents.

Also available are kits with the reagents useful for performing sequencing, fingerprinting, and mapping procedures. The kits will have various compartments with the desired necessary reagents, e.g., substrate, labeling reagents for target samples, buffers, and other useful accompanying products.

C. Mapping

The present invention also provides the means for mapping sequences within enormous stretches of sequence. For example, nucleotide sequences may be mapped within enormous chromosome size sequence maps. For example, it would be possible to map a chromosomal location within the chromosome which contains hundreds of millions of nucleotide base pairs. In addition, the mapping and fingerprinting embodiments allow for testing of chromosomal translocations, one of the standard problems for which amniocentesis is performed.

Thus, the present invention provides a powerful tool and the means for performing sequencing, fingerprinting, and mapping functions on polymers. Although most easily and directly applicable to polynucleotides, polypeptides, carbohydrates, and other sorts of molecules can be advantageously utilized using the present technology.

The present invention will be better understood by reference to the following illustrative examples. The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL

I. Sequencing
   A. polynucleotide
   B. polypeptide
   C. short peptide
      1. Herz antibody identification
II. Fingerprinting
   A. polynucleotide fingerprint
   B. peptide fingerprint
   C. cell classification scheme
   D. temporal development scheme
      1. developmental antigens
      2. developmental mRNA expression
   E. diagnostic test
      1. viral identification
      2. bacterial identification
      3. other microbiological identifications
      4. allergy test (immobilized antigens)
   F. individual (animal/plant) identification
      1. genetic
      2. immunological
   G. genetic screen
      1. test alleles with markers
      2. amniocentesis
III. Mapping
   A. positionally located clones (caged biotin)
      1. short probes, long targets
      2. long targets, short probes
   B. positionally defined clones
IV. Conclusion

* * * * * * * * * * * * * * *

Relevant applications whose techniques are incorporated herein by reference are Pirrung, et al., Serial No. 07/362,901, filed June 7, 1989, now abandoned; Pirrung et al. (1992) U.S. Pat. No. 5,143,854; Barrett, et al., Serial No. 07/435,316 filed November 13, 1989, now abandoned; Barrett, et al. (1993) U.S. Pat. No. 5,252,743; and commonly assigned and simultaneously filed applications Serial No. 07/624,120, now abandoned, and Serial No. 07/626,730.

Also, additional relevant techniques are described, e.g., in Sambrook, J., et al. (1989) Molecular Cloning: a Laboratory Manual, 2d Ed., vols 1-3, Cold Spring Harbor Press, New York; Greenstein and Winitz (1961) Chemistry of the Amino Acids, Wiley and Sons, New York; Bodzansky, M. (1988) Peptide Chemistry: a Practical Textbook, Springer-Verlag, New York;

Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Press, New York; Glover, D. (ed.) (1987) DNA Cloning: A Practical Approach, vols 1-3, IRL Press, Oxford; Bishop and Rawlings (1987) Nucleic Acid and Protein Sequence Analysis: A Practical Approach, IRL Press, Oxford; Hames and Higgins (1985) Nucleic Acid Hybridisation: A Practical Approach, IRL Press, Oxford; Wu et al. (1989) Recombinant DNA Methodology, Academic Press, San Diego; Goding (1986) Monoclonal Antibodies: Principles and Practice, (2d ed.), Academic Press, San Diego; Finegold and Barron (1986) Bailey and Scott's Diagnostic Microbiology, (7th ed.), Mosby Co., St. Louis; Collins et al. (1989) Microbiological Methods, (6th ed.), Butterworth, London; Chaplin and Kennedy (1986) Carbohydrate Analysis: A Practical Approach, IRL Press, Oxford; Van Dyke (ed.) (1985) Bioluminescence and Chemiluminescence: Instruments and Applications, vol 1, CRC Press, Boca Rotan; and Ausubel et al. (ed.) (1990) Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, New York; each of which is hereby incorporated herein by reference.

The following examples are provided to illustrate the efficacy of the inventions herein. All operations were conducted at about ambient temperatures and pressures unless indicated to the contrary.

I. SEQUENCING

A. Polynucleotide

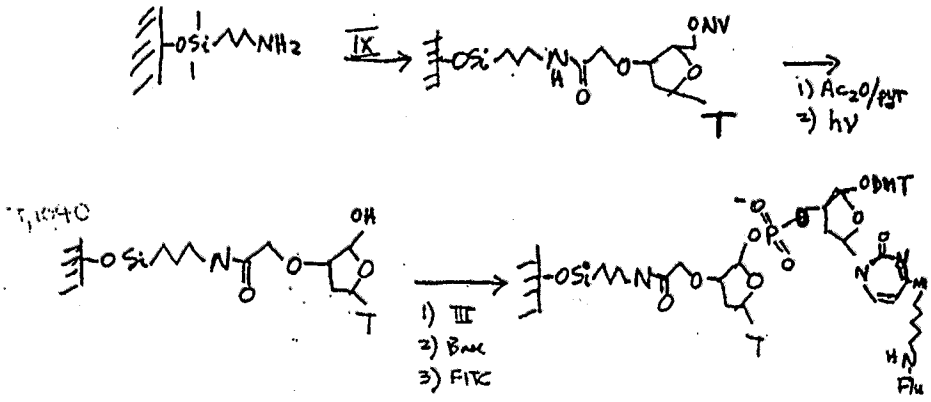
1. HPLC of the photolysis of 5'-O-nitroveratryl-thymidine.

In order to determine the time for photolysis of 5'-O-nitroveratryl thymidine to thymidine a 100 µM solution of NV-Thym-OH (5'-O-nitroveratryl thymidine) in dioxane was made and ~200 µl aliquots were irradiated (in a quartz cuvette 1 cm x 2 mm) at 362.3 nm for 20 sec, 40 sec, 60 sec, 2 min, 5 min, 10 min, 15 min, and 20 min. The resulting irradiated mixtures were then analyzed by HPLC using a Varian MicroPak SP column (C_{18} analytical) at a flow rate of 1 ml/min and a solvent system of 40% CH_3CN and 60% water. Thymidine has a retention time of

1.2 min and NVO-Thym-OH has a retention time of 2.1 min. It was seen that after 10 min of exposure the deprotection was complete.

2. Preparation and Detection of Thymidine-Cytidine dimer (FITC)

The reaction is illustrated:



To an aminopropylated glass slide (standard VLSIPS™ Technology) was added a mixture of the following:

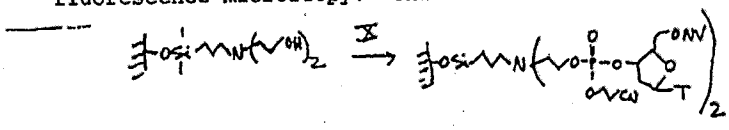
12.2 mg of NVO-Thym-CO_2H (IX)
3.4 mg of HOBT (N-hydroxybenztriazal)
8.8 µl DIEA (Diisopropylethylamine)
11.1 mg BOP reagent
2.5 ml DMF

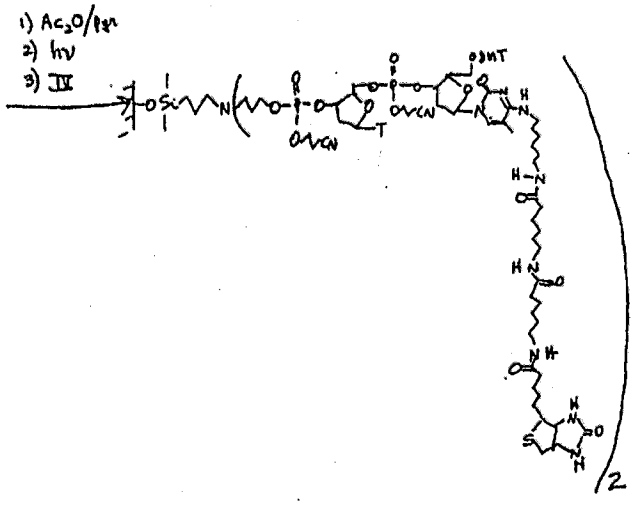
After 2 h coupling time (standard VLSIPS) the plate was washed, acetylated with acetic anhydride/pyridine, washed, dried, and photolyzed in dioxane at 362 nm at 14 mW/cm^2 for 10

min using a 500 µm checkerboard mask. The slide was then taken and treated with a mixture of the following:

107 mg of FMOC-amine modified C (III)
21 mg of tetrazole
1 ml anhydrous CH_3CN

After being treated for approximately 8 min, the slide was washed off with CH_3CN, dried, and oxidized with I_2/H_2O/THF/lutidine for 1 min. The slide was again washed, dried, and treated for 30 min with a 20% solution of DBU in DMF. After thorough rinsing of the slide, it was next exposed to a FITC solution (1mM fluorescein isothiocyanate [FITC] in DMF) for 50 min, then washed, dried, and examined by fluorescence microscopy. This reaction is illustrated:





3. Preparation and Detection of Thymidine-Cytidine dimer (Biotin)

An aminopropyl glass slide, was soaked in a solution of ethylene oxide (20% in DMF) to generate a hydroxylated surface. The slide was added to a mixture of the following:

32 mg of NVO-T-OCED (X)
11 mg of tetrazole
0.5 ml of anhydrous CH_3CN

After 8 min the plate was then rinsed with acetonitrile, then oxidized with I_2/H_2O/THF/lutidine for 1 min, washed and dried. The slide was then exposed to a 1:3 mixture of acetic anhydride:pyridine for 1 h, then washed and dried. The substrate was then photolyzed in dioxane at 362 nm at 14 mW/cm^2 for 10 min using a 500µm checkerboard mask, dried, and then treated with a mixture of the following:

65 mg of biotin modified C (IV)
11 mg of tetrazole
0.5 ml anhydrous CH_3CN

After 8 min the slide was washed with CH_3CN then oxidized with I_2/H_2O/THF/lutidine for 1 min, washed, and then dried. The slide was then soaked for 30 min in a PBS/0.05% Tween 20 buffer and the solution then shaken off. The slide was next treated with FITC-labeled streptavidin at 10 µg/ml in the same buffer system for 30 min. After this time the streptavidin-buffer system was rinsed off with fresh PBS/0.05% Tween 20 buffer and then the slide was finally agitated in distilled water for about 1/2 h. After drying, the slide was examined by fluorescence microscopy ~~(see Fig. 2 and Fig. 3)~~.

4. substrate preparation

Before attachment of reactive groups it is preferred to clean the substrate which is, in a preferred embodiment, a glass substrate such as a microscope slide or cover slip. A roughened surface will be useable but a plastic or other solid substrate is also appropriate. According to one embodiment the slide is soaked in an alkaline bath consisting of, e.g., 1 liter of 95% ethanol with 120 ml of water and 120 grams of sodium hydroxide for 12 hours. The slides are washed with a buffer and under running water, allowed to air dry, and rinsed with a solution of 95% ethanol.

The slides are then aminated with, e.g., aminopropyltriethoxysilane for the purpose of attaching amino groups to the glass surface on linker molecules, although other

omega functionalized silanes could also be used for this purpose. In one embodiment 0.1% aminopropyltriethoxysilane is utilized, although solutions with concentrations from 10^{-7}% to 10% may be used, with about 10^{-3}% to 2% preferred. A 0.1% mixture is prepared by adding to 100 ml of a 95% ethanol/5% water mixture, 100 microliters (µl) of aminopropyltriethoxysilane. The mixture is agitated at about ambient temperature on a rotary shaker for an appropriate amount of time, e.g., about 5 minutes. 500 µl of this mixture is then applied to the surface of one side of each cleaned slide. After 4 minutes or more, the slides are decanted of this solution and thoroughly rinsed three times or more by dipping in 100% ethanol.

After the slides dry, they are heated in a 110-120°C vacuum oven for about 20 minutes, and then allowed to cure at room temperature for about 12 hours in an argon environment. The slides are then dipped into DMF (dimethylformamide) solution, followed by a thorough washing with methylene chloride.

5. linker attachment, blocking of free sites

The aminated surface of the slide is then exposed to about 500 µl of, for example, a 30 millimolar (mM) solution of NVOC-nucleotide- NHS (N-hydroxysuccinimide) in DMF for attachment of a NVOC-nucleotide to each of the amino groups. See, e.g., SIGMA Chemical Company for various nucleotide derivatives. The surface is washed with, for example, DMF, methylene chloride, and ethanol.

Any unreacted aminopropyl silane on the surface, i.e., those amino groups which have not had the NVOC-nucleotide attached, are now capped with acetyl groups (to prevent further reaction) by exposure to a 1:3 mixture of acetic anhydride in pyridine for 1 hour. Other materials which may perform this residual capping function include trifluoroacetic anhydride, formicacetic anhydride, or other reactive acylating agents. Finally, the slides are washed again with DMF, methylene chloride, and ethanol.

6. synthesis of eight trimers of C and T

Fig. 24 illustrates a possible synthesis of the eight trimers of the two-monomer set: cytosine and thymine (represented by C and T, respectively). A glass slide bearing silane groups terminating in 6-nitroveratryloxycarboxamide (NVOC-NH) residues is prepared as a substrate. Active esters (pentafluorophenyl, OBt, etc.) of cytosine and thymine protected at the 5' hydroxyl group with NVOC are prepared as reagents. While not pertinent to this example, if side chain protecting groups are required for the monomer set, these must not be photoreactive at the wavelength of light used to protect the primary chain.

For a monomer set of size n, n x l cycles are required to synthesize all possible sequences of length l. A cycle consists of:

1. Irradiation through an appropriate mask to expose the 5'-OH groups at the sites where the next residue is to be added, with appropriate washes to remove the by-products of the deprotection.
2. Addition of a single activated and protected (with the same photochemically-removable group) monomer, which will react only at the sites addressed in step 1, with appropriate washes to remove the excess reagent from the surface.

The above cycle is repeated for each member of the monomer set until each location on the surface has been extended by one residue in one embodiment. In other embodiments, several residues are sequentially added at one location before moving on to the next location. Cycle times will generally be limited by the coupling reaction rate, now as short as about 10 min in automated oligonucleotide synthesizers. This step is optionally followed by addition of a protecting group to stabilize the array for later testing. For some types of polymers (e.g., peptides), a final deprotection of the entire surface (removal of photoprotective side chain groups) may be required.

More particularly, as shown in Fig. 2A, the glass 20 is provided with regions 22, 24, 26, 28, 30, 32, 34, and 36. Regions 30, 32, 34, and 36 are masked, indicated by the hatched regions, as shown in Fig. 2B and the glass is irradiated by the bright regions 22, 24, 26, and 28, and exposed to a reagent containing a photosensitive blocked C (e.g., cytosine derivative), with the resulting structure shown in Fig. 2C. The substrate is carefully washed and the reactants removed. Thereafter, regions 22, 24, 26, and 28 are masked, as indicated by the hatched region, the glass is irradiated (as shown in Fig. 2D), as indicated by the bright regions, at 30, 32, 34, and 36, and exposed to a photosensitive blocked reagent containing T (e.g., thymine derivative), with the resulting structure shown in Fig. 2E. The process proceeds, consecutively masking and exposing the sections as shown until the structure shown in Fig. 2M is obtained. The glass is irradiated and the terminal groups are, optionally, capped by acetylation. As shown, all possible trimers of cytosine/thymine are obtained.

In this example, no side chain protective group removal is necessary, as might be common in modified nucleotides. If it is desired, side chain deprotection may be accomplished by treatment with ethanedithiol and trifluoro-acetic acid.

In general, the number of steps needed to obtain a particular polymer chain is defined by:

$$n \times l \quad (1)$$

where:

n = the number of monomers in the basis set of monomers, and

l = the number of monomer units in a polymer chain.

Conversely, the synthesized number of sequences of length l will be:

$$n^l. \quad (2)$$

Of course, greater diversity is obtained by using masking strategies which will also include the synthesis of polymers having a length of less than l. If, in the extreme case, all polymers having a length less than or equal to l are

synthesized, the number of polymers synthesized will be:

$$n^1 + n^{1-1} + \ldots + n^1. \quad (3).$$

The maximum number of lithographic steps needed will generally be n for each "layer" of monomers, i.e., the total number of masks (and, therefore, the number of lithographic steps) needed will be n x l. The size of the transparent mask regions will vary in accordance with the area of the substrate available for synthesis and the number of sequences to be formed. In general, the size of the synthesis areas will be:

size of synthesis areas = (A)/(S)

where:

A is the total area available for synthesis; and
S is the number of sequences desired in the area.

It will be appreciated by those of skill in the art that the above method could readily be used to simultaneously produce thousands or millions of oligomers on a substrate using the photolithographic techniques disclosed herein. Consequently, the method results in the ability to practically test large numbers of, for example, di, tri, tetra, penta, hexa, hepta, octa, nona, deca, even dodecanucleotides, or larger polynucleotides (or correspondingly, polypeptides).

The above example has illustrated the method by way of a manual example. It will of course be appreciated that automated or semi-automated methods could be used. The substrate would be mounted in a flow cell for automated addition and removal of reagents, to minimize the volume of reagents needed, and to more carefully control reaction conditions. Successive masks will be applicable manually or automatically. See, e.g., Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and Serial No. 07/624,120, now abandoned.

7. labeling of target

The target oligonucleotide can be labeled using standard procedures referred to above. As discussed, for certain situations, a reagent which recognizes interaction, e.g., ethidium bromide, may be provided in the detection step. Alternatively, fluorescence labeling techniques may be

applied, see, e.g., Smith, et al. (1986) Nature, 321: 674-679; and Prober, et al. (1987) Science, 238:336-341. The techniques described therein will be followed with minimal modifications as appropriate for the label selected.

8. dimers of A, C, G, and T

The described technique may be applied, with photosensitive blocked nucleotides corresponding to adenine, cytosine, guanine, and thymine, to make combinations of polynucleotides consisting of each of the four different nucleotides. All 16 possible dimers would be made using a minor modification of the described method.

9. 10-mers of A, C, G, and T

The described technique for making dimers of A, C, G, and T may be further extended to make longer oligonucleotides. The automated system described, e.g., in Pirrung et al. (1992) U.S. Pat. No. 5,143,854, and Serial No. 07/624,120, now abandoned, can be adapted to make all possible 10-mers composed of the 4 nucleotides A, C, G, and T. The photosensitive, blocked nucleotide analogues have been described above, and would be readily adaptable to longer oligonucleotides.

10. specific recognition hybridization to 10-mers

The described hybridization conditions are directly applicable to the sequence specific recognition reagents attached to the substrate, produced as described immediately above. The 10-mers have an inherent property of hybridizing to a complementary sequence. For optimum discrimination between full matching and some mismatch, the conditions of hybridization should be carefully selected, as described above. Careful control of the conditions, and titration of parameters should be performed to determine the optimum collective conditions.

11. hybridization

Hybridization conditions are described in detail,

e.g., in Hames and Higgins (1985) Nucleic Acid Hybridisation; A Practical Approach; and the considerations for selecting particular conditions are described, e.g., in Wetmur and Davidson, (1988) J. Mol. Biol. 31:349-370, and Wood et al. (1985) Proc. Natl. Acad. Sci. USA 82:1585-1588. As described above, conditions are desired which can distinguish matching along the entire length of the probe from where there is one or more mismatched bases. The length of incubation and conditions will be similar, in many respects, to the hybridization conditions used in Southern blot transfers. Typically, the GC bias may be minimized by the introduction of appropriate concentrations of the alkylammonium buffers, as described above.

Titration of the temperature and other parameters is desired to determine the optimum conditions for specificity and distinguishability of absolutely matched hybridization from mismatched hybridization.

A fluorescently labeled target or set of targets are generated, as described in Prober, et al. (1987) Science 238:336-341, or Smith, et al. (1986) Nature 321:674-679. Preferably, the target or targets are of the same length as, or slightly longer, than the oligonucleotide probes attached to the substrate and they will have known sequences. Thus, only a few of the probes hybridize perfectly with the target, and which particular ones did would be known.

The substrate and probes are incubated under appropriate conditions for a sufficient period of time to allow hybridization to completion. The time is measured to determine when the probe-target hybridizations have reached completion. A salt buffer which minimizes GC bias is preferred, incorporating, e.g., buffer, such as tetramethyl ammonium or tetraethyl ammonium ion at between about 2.4 and 3.0 M. See Wood, et al. (1985) Proc. Nat'l Acad. Sci. USA 82:1585-1588. This time is typically at least about 30 min, and may be as long as about 1-5 days. Typically very long matches will hybridize more quickly, very short matches will hybridize less quickly, depending upon relative target and probe concentrations. The hybridization will be performed under

conditions where the reagents are stable for that time duration.

Upon maximal hybridization, the conditions for washing are titrated. Three parameters initially titrated are time, temperature, and cation concentration of the wash step. The matrix is scanned at various times to determine the conditions at which the distinguishability between true perfect hybrid and mismatched hybrid is optimized. These conditions will be preferred in the sequencing embodiments.

12. positional detection of specific interaction

As indicated above, the detection of specific interactions may be performed by detecting the positions where the labeled target sequences are attached. Where the label is a fluorescent label, the apparatus described, e.g., in Pirrung et al. (1992) U.S. Pat. No. 5,143,854; and Serial No. 07/624,120, now abandoned, may be advantageously applied. In particular, the synthetic processes described above will result in a matrix pattern of specific sequences attached to the substrate, and a known pattern of interactions can be converted to corresponding sequences.

In an alternative embodiment, a separate reagent which differentially interacts with the probe and interacted probe/targets can indicate where interaction occurs or does not occur. A single-strand specific reagent will indicate where no interaction has taken place, while a double-strand specific reagent will indicate where interaction has taken place. An intercalating dye, e.g., ethidium bromide, may be used to indicate the positions of specific interaction.

13. analysis

Conversion of the positional data into sequence specificity will provide the set of subsequences whose analysis by overlap segments, may be performed, as described above. Analysis is provided by the methodology described above, or using, e.g., software available from the Genetic Engineering Center, P.O. Box 794, 11000 Belgrade, Yugoslavia (Yugoslav

group). See, also, Macevicz, PCT publication no. WO 90/04652, which is hereby incorporated herein by reference.

B. Polypeptide

The description of the preparation of short peptides on a substrate incorporates by reference sections in Pirrung et al. (1992) U.S. Pat. No. 5,143,854, and described below.

1. slide preparation

Preparation of the substrate follows that described above for nucleotides.

2. linker attachment, blocking of free sites

The aminated surface of the slide is exposed to about 500 µl of, e.g., a 30 millimolar (mM) solution of NVOC-GABA (gamma amino butyric acid) NHS (N-hydroxysuccinimide) in DMF for attachment of a NVOC-GABA to each of the amino groups. The surface is washed with, for example, DMF, methylene chloride, and ethanol. See Serial No. 07,624,120, now abandoned, for details on amino acid chemistry.

Any unreacted aminopropyl silane on the surface, i.e., those amino groups which have not had the NVOC-GABA attached, are now capped with acetyl groups (to prevent further reaction) by exposure to a 1:3 mixture of acetic anhydride in pyridine for 1 hour. Other materials which may perform this residual capping function include trifluoroacetic anhydride, formicacetic anhydride, or other reactive acylating agents. Finally, the slides are washed again with DMF, methylene chloride, and ethanol.

3. synthesis of 8 trimers of "A" and "B"

See Pirrung et al. (1992) U.S. Pat. No. 5,143,854 which describes the preparation of glycine and phenylalanine trimers. The technique is similar to the method described above for making triners of C and T, but substituting photosensitive blocked glycine for the C derivative and photosensitive blocked phenylalamine for the T derivative.

4. synthesis of a dimer of an aminopropyl group and a fluorescent group

In synthesizing the dimer of an aminopropyl group and a fluorescent group, a functionalized Durapore™ membrane was used as a substrate. The Durapore™ membrane was a polyvinylidine difluoride with aminopropyl groups. The aminopropyl groups were protected with the DDZ group by reaction of the carbonyl chloride with the amino groups, a reaction readily known to those of skill in the art. The surface bearing these groups was placed in a solution of THF and contacted with a mask bearing a checkerboard pattern of 1 mm opaque and transparent regions. The mask was exposed to ultraviolet light having a wavelength down to at least about 280 nm for about 5 minutes at ambient temperature, although a wide range of exposure times and temperatures may be appropriate in various embodiments of the invention. For example, in one embodiment, an exposure time of between about 1 and 5000 seconds may be used at process temperatures of between -70 and +50°C.

In one preferred embodiment, exposure times of between about 1 and 500 seconds at about ambient pressure are used. In some preferred embodiments, pressure above ambient is used to prevent evaporation.

The surface of the membrane was then washed for about 1 hour with a fluorescent label which included an active ester bound to a chelate of a lanthanide. Wash times will vary over a wide range of values from about a few minutes to a few hours. These materials fluoresce in the red and the green visible region. After the reaction with the active ester in the fluorophore was complete, the locations in which the fluorophore was bound could be visualized by exposing them to ultraviolet light and observing the red and the green fluorescence. It was observed that the derivatized regions of the substrate closely corresponded to the original pattern of the mask.

5. demonstration of signal capability

Signal detection capability was demonstrated using a

low-level standard fluorescent bead kit manufactured by Flow Cytometry Standards and having model no. 824. This kit includes 5.8 μm diameter beads, each impregnated with a known number of fluorescein molecules.

One of the beads was placed in the illumination field on the scan stage in a field of a laser spot which was initially shuttered. After being positioned in the illumination field, the photon detection equipment was turned on. The laser beam was unblocked and it interacted with the particle bead, which then fluoresced. Fluorescence curves of beads impregnated with 7,000 and 29,000 fluorescein molecules, are shown in Figs. 11A and 11B, respectively of Pirrung et al. (1992) U.S. Pat. No. 5,143,854. On each curve, traces for beads without fluorescein molecules are also shown. These experiments were performed with 488 nm excitation, with 100 μW of laser power. The light was focused through a 40 power 0.75 NA objective.

The fluorescence intensity in all cases started off at a high value and then decreased exponentially. The fall-off in intensity is due to photobleaching of the fluorescein molecules. The traces of beads without fluorescein molecules are used for background subtraction. The difference in the initial exponential decay between labeled and nonlabeled beads is integrated to give the total number of photon counts, and this number is related to the number of molecules per bead. Therefore, it is possible to deduce the number of photons per fluorescein molecule that can be detected. This calculation indicates the radiation of about 40 to 50 photons per fluorescein molecule are detected.

6. determination of the number of molecules per unit area

Aminopropylated glass microscope slides prepared according to the methods discussed above were utilized in order to establish the density of labeling of the slides. The free amino termini of the slides were reacted with FITC (fluorescein isothiocyanate) which forms a covalent linkage with the amino group. The slide is then scanned to count the number of

fluorescent photons generated in a region which, using the estimated 40-50 photons per fluorescent molecule, enables the calculation of the number of molecules which are on the surface per unit area.

A slide with aminopropyl silane on its surface was immersed in a 1 mM solution of FITC in DMF for 1 hour at about ambient temperature. After reaction, the slide was washed twice with DMF and then washed with ethanol, water, and then ethanol again. It was then dried and stored in the dark until it was ready to be examined.

Through the use of curves similar to those shown in Fig. 11 of Pirrung et al. (1992) U.S. Pat. No. 5,143,854, and by integrating the fluorescent counts under the exponentially decaying signal, the number of free amino groups on the surface after derivitization was determined. It was determined that slides with labeling densities of 1 fluorescein per 10^3x10^3 to ~2x2 nm could be reproducibly made as the concentration of aminopropyltriethoxysilane varied from 10^{-3}% to 10^{-1}%.

7. removal of NVOC and attachment of a fluorescent marker

NVOC-GABA groups were attached as described above. The entire surface of one slide was exposed to light so as to expose a free amino group at the end of the gamma amino butyric acid. This slide, and a duplicate which was not exposed, were then exposed to fluorescein isothiocyanate (FITC).

Fig. 12A of Pirrung et al. (1992) U.S. Pat. No. 5,143,854 illustrates the slide which was not exposed to light, but which was exposed to FITC. The units of the x axis are time and the units of the y axis are counts. The trace contains a certain amount of background fluorescence. The duplicate slide was exposed to 350 nm broadband illumination for about 1 minute (12 mW/cm^2, ~350 nm illumination), washed and reacted with FITC. A large increase in the level of fluorescence is observed, which indicates photolysis has exposed a number of amino groups on the surface of the slides for attachment of a fluorescent marker.

8. use of a mask in removal of NVOC

The next experiment was performed with a 0.1% aminopropylated slide. Light from a Hg-Xe arc lamp was imaged onto the substrate through a laser-ablated chrome-on-glass mask in direct contact with the substrate.

This slide was illuminated for approximately 5 minutes, with 12 mW of 350 nm broadband light and then reacted with the 1 mM FITC solution. It was put on the laser detection scanning stage and a graph was plotted as a two-dimensional representation of position color-coded for fluorescence intensity. The experiment was repeated a number of times through various masks. The fluorescence patterns for a 100x100 μm mask, a 50 μm mask, a 20 μm mask, and a 10 μm mask indicate that the mask pattern is distinct down to at least about 10 μm squares using this lithographic technique.

9. attachment of YGGFL and subsequent exposure to herz antibody and goat anti-mouse antibody

In order to establish that receptors to a particular polypeptide sequence would bind to a surface-bound peptide and be detected, Leu enkephalin was coupled to the surface and recognized by an antibody. A slide was derivatized with 0.1% amino propyl-triethoxysilane and protected with NVOC. A 500 μm checkerboard mask was used to expose the slide in a flow cell using backside contact printing. The Leu enkephalin sequence (H_2N-tyrosine,glycine,glycine,phenylalanine,leucine-COOH, otherwise referred to herein as YGGFL) was attached via its carboxy end to the exposed amino groups on the surface of the slide. The peptide was added in DMF solution with the BOP/HOBT/DIEA coupling reagents and recirculated through the flow cell for 2 hours at room temperature.

A first antibody, known as the Herz antibody, was applied to the surface of the slide for 45 minutes at 2 μg/ml in a supercocktail (containing 1% BSA and 1% ovalbumin also in this case). A second antibody, goat anti-mouse fluorescein conjugate, was then added at 2 μg/ml in the supercocktail

buffer, and allowed to incubate for 2 hours.

The results of this experiment were plotted as fluorescence intensity as a function of position. This image was taken at 10 μm steps and showed that not only can deprotection be carried out in a well defined pattern, but also that (1) the method provided for successful coupling of peptides to the surface of the substrate, (2) the surface of a bound peptide was available for binding with an antibody, and (3) the detection apparatus capabilities were sufficient to detect binding of a receptor. Moreover, the Herz antibody is a sequence specific reagent which may be used advantageously as a sequence specific recognition reagent. It may be used, if specificity is high, for sequencing purposes, and, at least, for fingerprinting and mapping uses.

10. monomer-by-monomer formation of YGGFL and subsequent exposure to labeled antibody

Monomer-by-monomer synthesis of YGGFL and GGFL in alternate squares was performed on a slide in a checkerboard pattern and the resulting slide was exposed to the Herz antibody.

A slide is derivatized with the aminopropyl group, protected in this case with t-BOC (t-butoxycarbonyl). The slide was treated with TFA to remove the t-BOC protecting group. E-aminocaproic acid, which was t-BOC protected at its amino group, was then coupled onto the aminopropyl groups. The aminocaproic acid serves as a spacer between the aminopropyl group and the peptide to be synthesized. The amino end of the spacer was deprotected and coupled to NVOC-leucine. The entire slide was then illuminated with 12 mW of 325 nm broadband illumination. The slide was then coupled with NVOC-phenylalanine and washed. The entire slide was again illuminated, then coupled to NVOC-glycine and washed. The slide was again illuminated and coupled to NVOC-glycine to form the sequence shown in the last portion of Fig. 13A of Pirrung et al. (1992) U.S. Pat. No. 5,143,854.

Alternating regions of the slide were then

illuminated using a projection print using a 500x500 µm checkerboard mask; thus, the amino group of glycine was exposed only in the lighted areas. When the next coupling chemistry step was carried out, NVOC-tyrosine was added, and it coupled only at those spots which had received illumination. The entire slide was then illuminated to remove all the NVOC groups, leaving a checkerboard of YGGFL in the lighted areas and in the other areas, GGFL. The Herz antibody (which recognizes the YGGFL, but not GGFL) was then added, followed by goat anti-mouse fluorescein conjugate.

The resulting fluorescence scan showed dark areas containing the tetrapeptide GGFL, which is not recognized by the Herz antibody (and thus there is no binding of the goat anti-mouse antibody with fluorescein conjugate), and red areas in which YGGFL was present. The YGGFL pentapeptide is recognized by the Herz antibody and, therefore, there is antibody in the lighted regions for the fluorescein-conjugated goat anti-mouse to recognize.

Similar patterns for a 50 µm mask used in direct contact ("proximity print") with the substrate provided a pattern which was more distinct and the corners of the checkerboard pattern were touching as a result of the mask being placed in direct contact with the substrate (which reflects the increase in resolution using this technique).

11. monomer-by-monomer synthesis of YGGFL and PGGFL

A synthesis using a 50 µm checkerboard mask was conducted. However, P was added to the GGFL sites on the substrate through an additional coupling step. P was added by exposing protected GGFL to light through a mask, and subsequence exposure to P in the manner set forth above. Therefore, half of the regions on the substrate contained YGGFL and the remaining half contained PGGFL.

The fluorescence plot for this experiment showed the regions are again readily discernable between those in which binding did and did not occur. This experiment demonstrated that antibodies are able to recognize a specific sequence and

that the recognition is not length-dependent.

12. monomer-by-monomer synthesis of YGGFL and YPGGFL

In order to further demonstrate the operability of the invention, a 50 µm checkerboard pattern of alternating YGGFL and YPGGFL was synthesized on a substrate using techniques like those set forth above. The resulting fluorescence plot showed that the antibody was clearly able to recognize the YGGFL sequence and did not bind significantly at the YPGGFL regions.

13. synthesis of an array of sixteen different amino acid sequences and estimation of relative binding affinity to herz antibody

Using techniques similar to those set forth above, an array of 16 different amino acid sequences (replicated four times) was synthesized on each of two glass substrates. The sequences were synthesized by attaching the sequence NVOC-GFL across the entire surface of the slides. Using a series of masks, two layers of amino acids were then selectively applied to the substrate. Each region had dimensions of 0.25 cm x 0.0625 cm. The first slide contained amino acid sequences containing only L- amino acids while the second slide contained selected D- amino acids. Various regions on the first and second slides, were duplicated four times on each slide. The slides were then exposed to the Herz antibody and fluorescein-labeled goat anti-mouse antibodies.

A fluorescence plot of the first slide, which contained only L- amino acids showed red areas (indicating strong binding, i.e., 149,000 counts or more) and black areas (indicating little or no binding of the Herz antibody, i.e., 20,000 counts or less). The sequence YGGFL was clearly most strongly recognized. The sequences YAGFL and YSGFL also exhibited strong recognition of the antibody. By contrast, most of the remaining sequences showed little or no binding. The four duplicate portions of the slide were extremely consistent in the amount of binding shown therein.

A fluorescence plot of the D- amino acid slide indicated that strongest binding was exhibited by the YGGFL sequence. Significant binding was also detected to YaGFL, YsGFL, and YpGFL. The remaining sequences showed less binding with the antibody. Low binding efficiency of the sequence yGGFL was observed.

Table 6 lists the various sequences tested in order of relative fluorescence, which provides information regarding relative binding affinity.

Table 6
Apparent Binding to Herz Ab

L- a.a. Set	D- a.a. Set
YGGFL	YGGFL
YAGFL	YaGFL
YSGFL	YsGFL
LGGFL	YpGFL
FGGFL	fGGFL
YPGFL	yGGFL
LAGFL	faGFL
FAGFL	wGGFL
WGGFL	yaGFL
	fpGFL
	waGFL

14. illustrative alternative embodiment

According to an alternative embodiment of the invention, the methods provide for attaching to the surface a caged binding member which, in its caged form, has a relatively low affinity for other potentially binding species, such as receptors and specific binding substances. Such techniques are more fully described in copending application Serial No. 404,920, filed September 8, 1989, and incorporated herein by reference for all purposes. See also Serial No. 07/435,316, now abandoned, and Barrett et al. (1993) U.S. Pat. No. 5,252,743, each of which is hereby incorporated herein by reference.

According to this alternative embodiment, the invention provides methods for forming predefined regions on a surface of a solid support, wherein the predefined regions are capable of immobilizing receptors. The methods make use of caged binding members attached to the surface to enable selective activation of the predefined regions. The caged binding members are liberated to act as binding members ultimately capable of binding receptors upon selective activation of the predefined regions. The activated binding

members are then used to immobilize specific molecules such as receptors on the predefined region of the surface. The above procedure is repeated at the same or different sites on the surface so as to provide a surface prepared with a plurality of regions on the surface containing, for example, the same or different receptors. When receptors immobilized in this way have a differential affinity for one or more ligands, screenings and assays for the ligands can be conducted in the regions of the surface containing the receptors.

The alternative embodiment may make use of novel caged binding members attached to the substrate. Caged (unactivated) members have a relatively low affinity for receptors of substances that specifically bind to uncaged binding members when compared with the corresponding affinities of activated binding members. Thus, the binding members are protected from reaction until a suitable source of energy is applied to the regions of the surface desired to be activated. Upon application of a suitable energy source, the caging groups labilize, thereby presenting the activated binding member. A typical energy source will be light.

Once the binding members on the surface are activated they may be attached to a receptor. The receptor chosen may be a monoclonal antibody, a nucleic acid sequence, a drug receptor, etc. The receptor will usually, though not always, be prepared so as to permit attaching it, directly or indirectly, to a binding member. For example, a specific binding substance having a strong binding affinity for the binding member and a strong affinity for the receptor or a conjugate of the receptor may be used to act as a bridge between binding members and receptors if desired. The method uses a receptor prepared such that the receptor retains its activity toward a particular ligand.

Preferably, the caged binding member attached to the solid substrate will be a photoactivatable biotin complex, i.e., a biotin molecule that has been chemically modified with photoactivatable protecting groups so that it has a significantly reduced binding affinity for avidin or avidin analogs than does natural biotin. In a preferred embodiment,

the protecting groups localized in a predefined region of the surface will be removed upon application of a suitable source of radiation to give binding members, that is biotin or a functionally analogous compound having substantially the same binding affinity for avidin or avidin analogs as does biotin.

In another preferred embodiment, avidin or an avidin analog is incubated with activated binding members on the surface until the avidin binds strongly to the binding members. The avidin so immobilized on predefined regions of the surface can then be incubated with a desired receptor or conjugate of a desired receptor. The receptor will preferably be biotinylated, e.g., a biotinylated antibody, when avidin is immobilized on the predefined regions of the surface. Alternatively, a preferred embodiment will present an avidin/biotinylated receptor complex, which has been previously prepared, to activated binding members on the surface.

II. FINGERPRINTING

The above section on generation of reagents for sequencing provides specific reagents useful for fingerprinting applications. Fingerprinting embodiments may be applied towards polynucleotide fingerprinting, polypeptide fingerprinting, cell and tissue classification, cell and tissue temporal development stage classification, diagnostic tests, forensic uses for individual identification, classification of organisms, and genetic screening of individuals. Mapping applications are also described below.

A. Polynucleotide Fingerprint

Polynucleotide fingerprinting may use reagents similar to those described above for probing a sequence for the presence of specific subsequences found therein. Typically, the subsequences used for fingerprinting will be longer than the sequences used in oligonucleotide sequencing. In particular, specific long segments may be used to determine the similarity of different samples of nucleic acids. They may also be used to fingerprint whether specific combinations of information are provided therein. Particular probe sequences

are selected and attached in a positional manner to a substrate. The means for attachment may be either using a caged biotin method described, e.g., in Barrett et al. (1993) U.S. Pat. No. 5,242,743, or by another method using targeting molecules. For example, a short polypeptide of specific sequence may be attached to an oligonucleotide and targeted to specific positions on a substrate having antibodies attached thereto, the antibodies exhibiting specificity for binding to those short peptide sequences. In another embodiment, an unnatural nucleotide or similar complementary binding molecule may be attached to the fingerprinting probe and the probe thereby directed towards complementary sequences on a VLSIPS substrate. Typically, unnatural nucleotides would be preferred, e.g., unnatural optical isomers, which would not interfere with natural nucleotide interactions.

Having produced a substrate with particular fingerprint probes attached thereto at positionally defined regions, the substrate may be used in a manner quite similar to the sequencing embodiment to provide information as to whether the fingerprint probes are detecting the corresponding sequence in a target sequence. This will often provide information similar to a Southern blot hybridization.

B. Polypeptide Fingerprint

A polypeptide fingerprint may be performed using antibodies which recognize specific antigens on the polypeptide. For example, monoclonal antibodies which recognize specific sequences or antigens on a polypeptide may be used to determine whether those epitopes are found on a particular protein. For example, particular patterns of epitopes would be found on various types of proteins. This will lead to the discovery that specific epitopes, or antigenic determinants, which are characteristic of, e.g., beta sheet segments, will be identified as will particular different types of domains in various protein types. Thus, a screening method may be devised which can classify polypeptides, either native or denatured, into various new classes defined by the epitopes existing thereon.

136

In addition, once the substrate is generated in the manners described above, a target peptide is exposed to the substrate. The target may be either native or denatured, though the conditions used to denature the polypeptide may interfere with the specific interaction between the polypeptide and the recognition reagent. This method is not dependent on the fact that the polypeptide is a single chain, thus protein complexes may also be fingerprinted using this methodology. Structures such as multi-subunit proteins, associations of proteins, ribosomes, nucleosomes, and other small cellular structures may also be fingerprinted and classified according to the presence of specific recognizable features thereon.

Peptide fingerprinting may be useful, for example, in correlating with particular physiological conditions or developmental stages of a cell or organism. Thus, a biological sample may be fingerprinted to determine the presence in that sample of a plurality of different polypeptides which are each individually fingerprinted. In an alternative embodiment, a polypeptide itself is not fingerprinted but a biological sample is fingerprinted searching for specific epitopes, e.g., polypeptide, carbohydrate, nucleic acid, or any of a number of other specific recognizable structural features.

The conditions for the interactions using antibodies is described, e.g., in Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Press, New York. The conditions should be titrated for temperature, buffer composition, time, and other important parameters in an antibody interaction.

C. Cell Classification Scheme

The present invention can be used for cell classification using fingerprinting type technology as described above in the polypeptide fingerprint. Classes of cells are typically defined by the presence of common functions which are usually reflected by structural features. Thus, a plant cell is classified differently from an animal cell by a number of structural features. Given an unknown cell, the present invention provides improved means for distinguishing

the different cell types. Once a cell classification scheme is developed and the structural features which define it are identified using the present invention, homogeneous cell population expressing these features may be separated from others. Standard cell sorters may be coupled with recognition reagents and labels which can distinguish various cell types.

a. T-Cell Classes

T-cell classes are defined on the basis of expression of particular antigens characteristic of each class. For example, mouse T-cell differentiation markers include the LY antigens. With the plurality of different antigens which may be tested using antibody or other recognition reagents, new populations and classes of cells may be defined. For example, different neural cell types may be defined on the basis of cell surface antigens. Different tissue types will be defined on the basis of tissue specific antigens. Developmental cell classes will be similarly defined. All of these screenings can make use of the VLSIPS substrates with specific recognition molecules attached thereto. The substrates are exposed to the cell types directly, assaying for attachment of cells to specific regions, or are exposed to products of a population of cells, e.g., a supernatant, or a cell lysate.

Once a cell classification scheme has been correlated with specific structural markers therein, reagents which recognize those features may be developed and used in a fluorescence activated cell sorter as described, e.g., in Dangl, J. and Herzenberg (1982) J. Immunological Methods 52: 1-14; and Becton Dickinson, Fluorescence Activated Cell Sorters Division, San Jose, California. This will provide a homogeneous population of cells whose function has been defined by structure.

b. B-Cell Classes

The present cell classification scheme may also be used to determine specific B-cell classes. For example, B-cells specific for producing IgM, IgG, IgD, IgE, and IgA may be defined by the internal expression of specific mRNA sequences

encoding each type of immunoglobulin. The classification scheme may depend on either extracellularly expressed markers which are correlated as being diagnostic of specific stages in development, or intracellular mRNA sequences which indicate particular functions.

D. Temporal Development Scheme

1. Developmental Antigens

The present fingerprinting invention also allows cell classification by expression of developmental antigens. For example, a lymphocyte stem cell expresses a particular combination of antigens. As the lymphocyte develops through a program developmental scheme, at various stages it expresses particular antigens which are diagnostic of particular stages in development. Again, the fingerprinting methodology allows for the definition of specific structural features which are diagnostic of developmental or functional features which will allow classification of cells into temporal developmental classes. Cells, products of those cells, or lysates of those cells will be assayed to determine the developmental stage of the source cells. In this manner, once a developmental stage is defined, specific synchronized populations of cells will be selected out of another population. These synchronized populations may be very important in determining the biological mechanisms of development.

2. Developmental mRNA Expression

Besides expressed antigens, the present invention also allows for fingerprinting of the mRNA population of a cell. In this fashion, the mRNA population, which should be a good determinant of developmental stage, will be correlated with other structural features of the cell. In this manner, cells at specific developmental stages will be characterized by the intracellular environment, as well as the extracellular environment. The present invention also allows the combination of definitions based, in part, upon antigens and, in part, upon mRNA expression.

In one embodiment, the two may be combined in a

single incubation step. A particular incubation condition may be found which is compatible with both hybridization recognition and non-hybridization recognition molecules. Thus, e.g., an incubation condition may be selected which allows both specificity of antibody binding and specificity of nucleic acid hybridization. This allows simultaneous performance of both types of interactions on a single matrix. Again, where developmental mRNA patterns are correlated with structural features, or with probes which are able to hybridize to intracellular mRNA populations, a cell sorter may be used to sort specifically those cells having desired mRNA population patterns.

E. Diagnostic Tests

The present invention also provides the ability to perform diagnostic tests. Diagnostic tests typically are based upon a fingerprint type assay, which tests for the presence of specific diagnostic structural features. Thus, the present invention provides means for viral strain identification, bacterial strain identification, and other diagnostic tests using positionally defined specific reagents. The present invention also allows for determining a spectrum of allergies, diagnosing a biological sample for any or all of the above, and testing for many other conditions.

1. Viral Identification

The present invention provides reagents and methodology for identifying viral strains. The specific reagents may be either antibodies or recognition proteins which bind to specific viral epitopes preferably surface exposed, but may make use of internal epitopes, e.g., in a denatured viral sample. In an alternative embodiment, the viral genome may be probed for specific sequences which are characteristic of particular viral strains. As above, a combination of the two may be performed simultaneously in a single interaction step, or in separate tests, e.g., for both genetic characteristics and epitope characteristics.

2. Bacterial Identification

Similar techniques will be applicable to identifying a bacterial source. This may be useful in diagnosing bacterial infections, or in classifying sources of particular bacterial species. For example, the bacterial assay may be useful in determining the natural range of survivability of particular strains of bacteria across regions of the country or in different ecological niches.

3. Other Microbiological Identifications

The present invention provides means for diagnosis of other microbiological and other species, e.g., protozoal species and parasitic species in a biological sample, but also provides the means for assaying a combination of different infections. For example, a biological specimen may be assayed for the presence of any or all of these microbiological species. In human diagnostic uses, typical samples will be blood, sputum, stool, urine, or other samples.

4. Allergy Tests

An immobilized set of antigens may be attached to a solid substrate and, instead of the standard skin reaction tests, a blood sample may be assayed on such a substrate to determine the presence of antibodies, e.g., IgE or other type antibodies, which may be diagnostic of an allergic or immunological susceptibility. A standard radioallergosorbent test (RAST) may be used to check a much larger population of antigens.

In addition, an allergy like test may be used to diagnose the immunological history of a particular individual. For example, by testing the circulating antibodies in a blood sample, which reflects the immunological history and memory of an individual, it may be determined what infections may not have been historically presented to the immune system. In this manner, it may be possible to specifically supplement an immune system for a short period of time with IgG fractions made up of specific types of gamma globulins. Thus, hepatitis gamma globulin injections may be better designed for a particular environment to which a person is expected to be exposed. This

also provides the ability to identify genetically equivalent individuals who have immunologically different experiences. Thus, a blood sample from an individual who has a particular combination of circulating antibodies will likely be different from the combination of circulating antibodies found in a genetically similar or identical individual. This could allow for the distinction between clones of particular animals, e.g., mice, rats, or other animals.

F. Individual Identification

The present invention provides the ability to fingerprint and identify a genetic individual. This individual may be a bacterial or lower microorganism, as described above in diagnostic tests, or of a plant or animal. An individual may be identified genetically or immunologically, as described.

1. Genetic

Genetic fingerprinting has been utilized in comparing different related species in Southern hybridization blots. Genetic fingerprinting has also been used in forensic studies, see, e.g., Morris et al. (1989) J. Forensic Science 34: 1311-1317, and references cited therein. As described above, an individual may be identified genetically by a sufficiently large number of probes. The likelihood that another individual would have an identical pattern over a sufficiently large number of probes may be statistically negligible. However, it is often quite important that a large number of probes be used where the statistical probability of matching is desired to be particularly low. In fact, the probes will optimally be selected for having high heterogeneity among the population. In addition, the fingerprint method may make use of the pattern of homologies indicated by a series of more and more stringent washes. Then, each position has both a sequence specificity and a homology measurement, the combination of which greatly increases the number of dimensions and the statistical likelihood of a perfect pattern match with another genetic individual.

2. Immunological

As indicated above in the diagnostic tests, it is possible to identify a particular immune system within a genetically homogeneous class of organisms by virtue of their immunological history. For example, a large colony of cloned mice may be distinguishable by virtue of each immunological history. For example, one mouse may have had an immunological response to exposure to antigen A to which her genetically identical sibling may have not been exposed. By virtue of this differential history, the first of the pair will likely have a high antibody titer against the antigen A whereas her genetically identical sibling will have not had a response to that antigen by virtue of never having been exposed to it. For this reason, immune systems may be identified by their immunological memories. Thus, immunological experience may also be a means for identifying a particular individual at a particular moment in her lifetime.

This same immunological screening may be used for other sorts of identifiable biological products. For example, an individual may be identified by her combination of expressed proteins. These proteins may reflect a physiological state of the individual, and would thus be useful in certain circumstances where diagnostic tests may be performed. For example, an individual may be identified, in part, by the presence of particular metabolic products.

In fact, a plant origin may be determined by virtue of having within its genome an unnatural sequence introduced to it by genetic breeders. Thus, a marker nucleic acid sequence may be introduced as a means to determine whether a genetic strain of a plant or animal originated from another particular source.

G. Genetic Screening

1. test alleles with markers

The present invention provides for the ability to screen for genetic variations of individuals. For example, a number of genetic diseases are linked with specific alleles. See, e.g., Scriber, C. et al. (eds.) (1989) The Metabolic Bases of Inherited Disease, McGraw-Hill, New York. In one

embodiment, cystic fibrosis has been correlated with a specific gene, see, Gregory et al. (1990) Nature 347: 382-386. A number of alleles are correlated with specific genetic deficiencies. See, e.g., McKusick, V. (1990) Genetic Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes, Johns Hopkins University Press, Baltimore; Ott, J. (1985) Analysis of Human Genetic Linkage, Johns Hopkins University Press, Baltimore; Track, R. et al. (1989) Banbury Report 32: DNA Technology and Forensic Science, Cold Spring Harbor Press, New York; each of which is hereby incorporated herein by reference.

2. Amniocentesis

Typically, amniocentesis is used to determine whether chromosome translocations have occurred. The mapping procedure may provide the means for determining whether these translocations have occurred, and for detecting particular alleles of various markers.

III. MAPPING

A. Positionally Located Clones

The present invention allows for the positional location of specific clones useful for mapping. For example, caged biotin may be used for specifically positioning a probe to a location on a matrix pattern.

In addition, the specific probes may be positionally directed to specific locations on a substrate by targeting. For example, polypeptide specific recognition reagents may be attached to oligonucleotide sequences which can be complementarily targeted to specific locations on a VLSIPS™ Technology substrate. Hybridization conditions, as applied for oligonucleotide probes, will be used to target the reagents to locations on a substrate having complementary oligonucleotides synthesized thereon. In another embodiment, oligonucleotide probes may be attached to specific polypeptide targeting reagents such as an antigen or antibody. These reagents can be directed towards a complementary antigen or antibody already attached to a VLSIPS substrate.

In another embodiment, an unnatural nucleotide which does not interfere with natural nucleotide complementary hybridization may be used to target oligonucleotides to particular positions on a substrate. Unnatural optical isomers of natural nucleotides should be ideal candidates.

In this way, short probes may be used to determine the mapping of long targets or long targets may be used to map the position of shorter probes. See, e.g., Craig et al. 1990 Nuc. Acids Res. 18: 2653-2660.

B. Positionally Defined Clones

Positionally defined clones may be transferred to a new substrate by either physical transfer or by synthetic means. Synthetic means may involve either a production of the probe on the substrate using the VLSIPS™ Technology synthetic methods, or may involve the attachment of a targeting sequence made by VLSIPS synthetic methods which will target that positionally defined clone to a position on a new substrate. Both methods will provide a substrate having a number of positionally defined probes useful in mapping.

IX. Conclusion

The present inventions provide greatly improved methods and apparatus for synthesis of polymers on substrates. It is to be understood that the above description is intended to be illustrative and not restrictive. Many embodiments will be apparent to those of skill in the art upon reviewing the above description. By way of example, the invention has been described primarily with reference to the use of photoremovable protective groups, but it will be readily recognized by those of skill in the art that sources of radiation other than light could also be used. For example, in some embodiments it may be desirable to use protective groups which are sensitive to electron beam irradiation, x-ray irradiation, in combination with electron beam lithograph, or x-ray lithography techniques. Alternatively, the group could be removed by exposure to an electric current. The scope of the invention should, therefore, be determined not with reference to the above

description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled.

All publications and patent applications referred to herein are incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually incorporated by reference. The present invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

WHAT IS CLAIMED IS:

1. A composition comprising a plurality of positionally distinguishable sequence specific reagents attached to a solid substrate, which reagents are capable of specifically binding to a predetermined subunit sequence of a preselected multi-subunit length having at least three subunits, said reagents representing substantially all possible sequences of said preselected length.

2. A composition of Claim 1, wherein said subunit sequence is a polynucleotide or a polypeptide.

3. A composition of Claim 1, wherein said preselected multi-subunit length is five subunits and said subunit sequence is a polynucleotide sequence.

4. A composition of Claim 1, wherein said specific reagent is an oligonucleotide of at least about five nucleotides.

5. A composition of Claim 1, wherein said specific reagent is a monoclonal antibody.

6. A composition of Claim 1, wherein said specific reagents are all attached to a single solid substrate.

7. A composition of Claim 1, wherein said reagents comprise about 3000 different sequences.

8. A composition of Claim 1, wherein said reagents represents at least about 25% of the possible subsequences of said preselected length.

9. A composition of Claim 1, wherein said reagents are localized in regions of the substrate having a density of at least 25 regions per square centimeter.

10. A composition of Claim 6, wherein said substrate has a surface area of less than about 4 square centimeters.

11. A method of analyzing a sequence of a polynucleotide or a polypeptide, said method comprising the step of:

a) exposing said polynucleotide or polypeptide to a composition of Claim 1.

12. A method of identifying or comparing a target sequence with a reference, said method comprising the step of:

a) exposing said target sequence to a composition of Claim 1;

b) determining the pattern of positions of said reagents which specifically interact with said target sequence; and

c) comparing said pattern with the pattern exhibited by said reference when exposed to said composition.

13. A method for sequencing a segment of a polynucleotide comprising the steps of:

a) combining:

i) a substrate comprising a plurality of chemically synthesized and positionally distinguishable oligonucleotides capable of recognizing defined oligonucleotide sequences; and

ii) a target polynucleotide; thereby forming high fidelity matched duplex structures of complementary subsequences of known sequence; and

b) determining which of said reagents have specifically interacted with subsequences in said target polynucleotide.

14. A method of Claim 13, wherein said segment is substantially the entire length of said polynucleotide.

15. A method for sequencing a polymer, said method comprising the steps of:

a) preparing a plurality of reagents which each specifically bind to a subsequence of preselected length;

b) positionally attaching each of said reagents to one or more solid phase substrates, thereby producing substrates of positionally definable sequence specific

probes;

c) combining said substrates with a target polymer whose sequence is to be determined; and

d) determining which of said reagents have specifically interacted with subsequences in said target polymer.

16. A method of Claim 15, wherein said substrates are beads.

17. A method of Claim 15, wherein said plurality of reagents comprise substantially all possible subsequences of said preselected length found in said target.

18. A method of Claim 15, wherein said solid phase substrates are a single substrate having attached thereto reagents recognizing substantially all possible subsequences of preselected length found in said target.

19. A method of Claim 15, further comprising the step of analyzing a plurality of said recognized subsequences to assemble a sequence of said target polymer.

20. A method of Claim 16, wherein at least some of said plurality of substrates have one subsequence specific reagent attached thereto, and said substrates are coded to indicate the specificity of said reagent.

21. A method of using a fluorescent nucleotide to

detect interactions with oligonucleotide probes of known sequence, said method comprising:

a) attaching said nucleotide to a target unknown polynucleotide sequence, and

b) exposing said target polynucleotide sequence to a collection of positionally defined oligonucleotide probes of known sequences to determine the sequences of said probes which interact with said target.

22. A method of Claim 21, further comprising the step of:

a) collating said known sequences to determine the overlaps of said known sequences to determine the sequence of said target sequence.

23. A method of mapping a plurality of sequences relative to one another, said method comprising:

a) preparing a substrate having a plurality of positionally attached sequence specific probes;

b) exposing each of said sequences to said substrate, thereby determining the patterns of interaction between said sequence specific probes and said sequences; and

c) determining the relative locations of said sequence specific probe interactions on

said sequences to determine the overlaps and order of said sequences.

24. A method of Claim 23, wherein said sequence specific probes are oligonucleotides.

25. A method of Claim 23, wherein said sequences are nucleic acid sequences.

PRODUCTS FOR DETECTING NUCLEIC ACIDS

ABSTRACT

The present invention provides methods and apparatus for sequencing, fingerprinting and mapping biological macromolecules, typically biological polymers. The methods make use of a plurality of sequence specific recognition reagents which can also be used for classification of biological samples, and to characterize their sources.

CERTIFICATE OF SERVICE

I hereby certify that on the 17th day of April, 2006, I caused to be electronically filed the foregoing document, **PUBLIC VERSION OF APPENDIX TO ILLUMINA, INC.'S OPENING MARKMAN BRIEF – VOLUME 1 OF 3,** with the Clerk of the Court using CM/ECF which will send notification of such filing to the following:

Jack B. Blumenfeld, Esq.
Mary Ellen Noreika, Esq.
Morris Nichols Arscht & Tunnell
1201 Market Street
Wilmington, DE 19801

Additionally, I hereby certify that on the 17th day of April, 2006, the foregoing document was served on the following via email:

Jack B. Blumenfeld, Esq.
Mary Ellen Noreika, Esq.
Morris Nichols Arscht & Tunnell
1201 Market Street
Wilmington, DE 19801

Daniel R. Reed, Esq.
Affymetrix, Inc.
6550 Vallejo Street, Suite 100
Emeryville, CA 94608
510.428.8500
Fax 510.428.8583

/s/ Richard K. Herrmann
Richard K. Herrmann (#405)
MORRIS, JAMES, HITCHENS
& WILLIAMS LLP
PNC Bank Center, 10th Floor
222 Delaware Avenue
P.O. Box 2306
Wilmington, Delaware 19899-2306
(302) 888-6800
rherrmann@morrisjames.com

Attorneys for Defendant
ILLUMINA, INC.